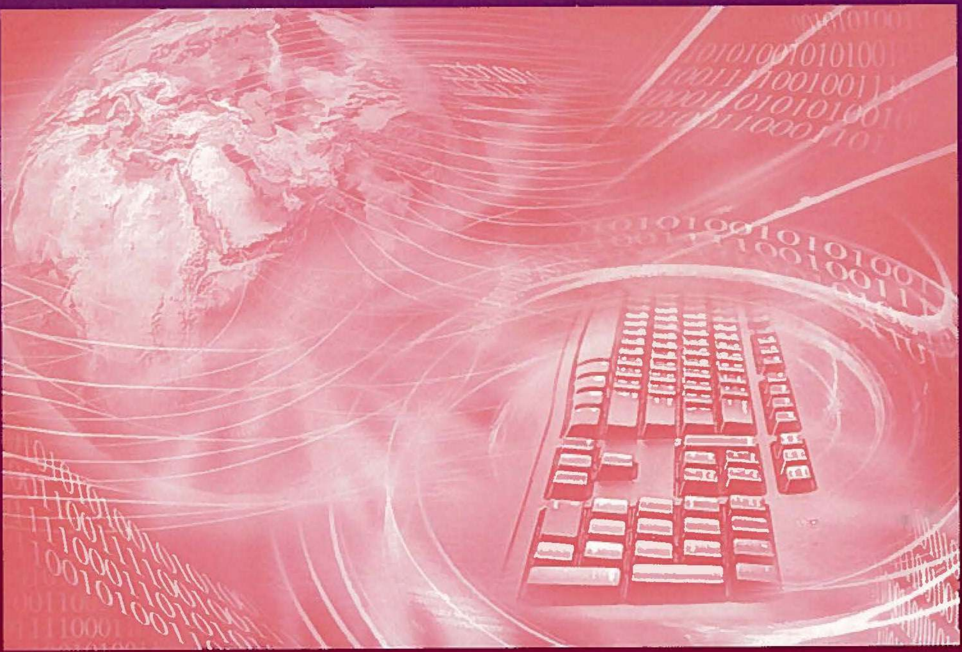
CHAPMAN & HALL/CRC COMPUTER and INFORMATION SCIENCE SERIES

Handbook of Energy-Aware and Green Computing

Volume 2



Edited by

Ishfaq Ahmad
Sanjay Ranka

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

41

Environmentally Opportunistic Computing

Paul Brenner
University of Notre Dame

AQ1

Douglas Thain
Aimee Buccellato
David B. Go

41.1 Introduction

Energy utilization by high-performance computing and information communications technology (HPC/ICT) is a critical resource management issue. In the United States, billions of dollars are spent annually on power and cool data systems. The 2007 U.S. Environmental Protection Agency "Report to Congress on Server and Data Center Efficiency" estimated that the United States spent $4.5 billion on electrical power to operate and cool HPC and ICT servers in 2006 with the same report forecasting that our national ICT electrical energy expenditure will nearly double—ballooning to $7.4 billion in 2011. Current energy demand for HPC/ICT is already 3% of U.S. electricity consumption and places considerable pressure on the domestic power grid: the peak load from HPC/ICT is estimated at 7 GW or the equivalent output of 15 base load power plants [1].

Recognizing that power resources for data centers are finite, several professional entities within the technology industry have begun to explore this problem including the High-Performance Buildings for High Tech Industries Team at Lawrence Berkley National Laboratory [2], the ASHRAE Technical Committee 9.9 for Mission Critical Facilities, Technology Spaces, and Electronic Equipment [3,4], the Uptime Institute [5], and the Green Grid [6]. At the same time, efforts by corporations, universities, and government labs to reduce their environmental footprint and more effectively manage their energy consumption have resulted in the development of novel waste heat exhaust and free cooling applications, such as the installation of the Barcelona Supercomputing Center, MareNostrum, in an eighteenth century Gothic masonry church [7], and novel waste heat recirculation applications, such as a centralized data center in Winnipeg that uses recirculated waste heat to heat the editorial offices of a newspaper directly above [8]. Similar centralized data centers in Israel [9] and Paris [10] use recaptured waste heat to condition adjacent office spaces and an on-site arboretum, respectively.

The ICT industry must continually evolve with the rapid development of individual but interrelated components which form the basis of internal infrastructure and end-user interfaces. Simultaneous changes in hardware, software, and networking technology are simply one example of the daily dynamic systems considerations. This high degree of variability and uncertainty makes it difficult to justify long-term infrastructure investments based on speculative total cost of ownership (TCO) analyses. It is not surprising therefore that major energy efficiency improvements to ICT data center facilities are often delayed until they can be coupled with renovations or new construction where improved reliability and capacity are key components of the funding justification. It is also worth noting that energy efficiency may be a smaller component of a total energy cost savings justification when a new facility location is selected based on proximity to less expensive power.

Given the technical- and market-based challenges to more energy efficient ICT facilities, it is promising to see publicly reported power usage effectiveness (PUE) values continuing to shrink toward 1.0. With hardware vendors and large ICT market players continuing to expand the acceptable ranges for free cooling, corporations and professional societies partnering to improve efficiency through cooperatives
AQ2 such as the Green Grid, and the U.S. federal government leading multiple initiatives through the DOE
AQ3 (FEMP Data Center Energy Efficiency) and EPA, it is plausible to expect continued (albeit slow) progress toward lower PUE values approaching 1.0. This begs the question our research intends to answer "Is a PUE of 1.0 the best possible outcome?"

41.2 Environmentally Opportunistic Computing Methodologies

Environmentally opportunistic computing (EOC) is the philosophy that ICT data centers do not need to be isolated entities, but can be integrated with existing buildings and facilities to synergistically share energy needs. Waste heat from high-output ICT can be harvested by the facility (waste heat recovery), and the facility's existing plumbing and air movement can be used as cooling for the ICT (free cooling). To that end, EOC uses *distributed* ICT infrastructure to create heat where it is already needed, to exploit cooling where it is already available, to utilize energy when and where it is least expensive, and to minimize the overall energy consumption of an organization. EOC engages sustainable computing at the macroscale, taking advantage of current scheduler, operating system, and hardware level efficiency [11–16] improvements as well as consolidation efforts in the virtualized data center. The aggressive growth of users—and the capability demanded by those users—must necessarily be met with new, integrated design paradigms, and EOC capitalizes on the dynamic mobility of virtualized services to exploit energy volatility for cost savings and lower environmental impact.

The focus of EOC research is to develop models, methods of delivery, and building/system design integrations that reach beyond current waste heat utilization applications and minimum energy standards to optimize the consumption of computational waste heat in the built environment. Transforming current approaches to ICT data center deployment on a wide scale via EOC requires the development of a systematic method for assessing, balancing, and effectively integrating various interrelated "market" forces (Table 41.1) related to the generation and efficient consumption of computer heat.

At the building scale, the efficient consumption of computer waste heat must be closely coordinated with building heating, ventilation, and air conditioning (HVAC) systems, whether these are existing technologies or new recovery and distribution systems designed specifically for waste heat recovery and free cooling. A sensor–control relationship must be established between these systems, the hardware they monitor, and the local input and output temperatures necessitated by the hardware and demanded by the building occupants, respectively. The controls network must mediate not only the dynamic relationship between source and target but also the variation in source and target interaction due to governing outside factors such as seasonal variations. In the colder winter months, the computational heat source can provide necessary thermal energy whereas the relationship inverts during the hot summer months when the facility can provide reasonably cool exhaust/makeup to the computational components.

TABLE 41.1 Relevant Market Forces for Integrating HPC/ICT into the Built Environment

1. User demand for computational capability
    a. Iterative examination of utilization patterns for various applications (science, business, entertainment, education, etc.)
    b. Iterative correlation of utilization characteristics with developing software, hardware, and network capabilities
2. Computational capability mobility and associated security concerns
    a. Evolution and adoption of grid/cloud computing and virtualization technology
    b. Security algorithms and implementations to allow sensitive/classified information transfer
3. Hardware thermal and environmental limits (temperature, humidity, particulate, etc.)
4. Facility concerns
    a. Integration with existing or novel active HVAC and/or passive systems
    b. General thermal performance variables (building materials, orientation, size, location of openings, etc.)
5. Facility occupant demands/concerns
    a. Thermal control (minimum user expectations and current standards and guidelines)
    b. Indoor air/environmental quality and perception of heat source (radiant computer heat)
6. Temperature variability (indoor/outdoor; day/night; seasonal)
7. Return on investment, TCO, and carbon reduction cost benefits/avoidance

The development of efficiency standards and increased expectations with respect to building occupant comfort require that the optimized integration of computational waste heat in a facility or group of facilities take into account the prevailing thermal comfort standards, like ASHRAE Standard 55-2004 Thermal Comfort Conditions for Human Occupancy [17] which specifies “the combinations of indoor space environment and personal factors that will produce thermal environmental conditions acceptable to 80% or more of the occupants within a space”; and more recent provisions for enhanced controllability of systems by building occupants, like the USGBC’s LEED rating system Environmental Quality Credit 6.2 Controllability of Systems [18] which calls for the provision of “individual comfort controls for a minimum of 50% of the building occupants to enable adjustments to suit individual task
AQ4 needs and preferences.” Comfort system control may be achieved as long as the building occupants have control over at least one of the primary indoor space environment criteria designated in ASHRAE Standard 55-2004: air temperature, radiant temperature, humidity, and air speed (USGBC 2007), all of which are critical considerations for the utilization and optimization of waste heat in a user-occupied facility.

41.3 EOC in Practice

The authors have evolved toward the EOC concept based a number of related smaller scale research
prototypes. Our first model framework called grid heating (GC) [19] specifically focused on utilization AQ5
and control of server exhaust heat within individual controlled human work centers. We were able to successfully demonstrate the dynamic, energy-based migration of ICT services in response to environmental stimuli [20]. The GC work proposed larger scale container-based possibilities as shown in Figure 41.1a. We then investigated CPU core level energy utilization characteristics for benchmark grid loads to shape our macroscale migration policies [21], finding dominant efficiency benefits of hibernation states over voltage scaling and disabling individual cores.

Most recently, the authors have constructed a heterogeneous, geographically distributed, multi-institutional grid infrastructure for executing service migration in production ICT environments and

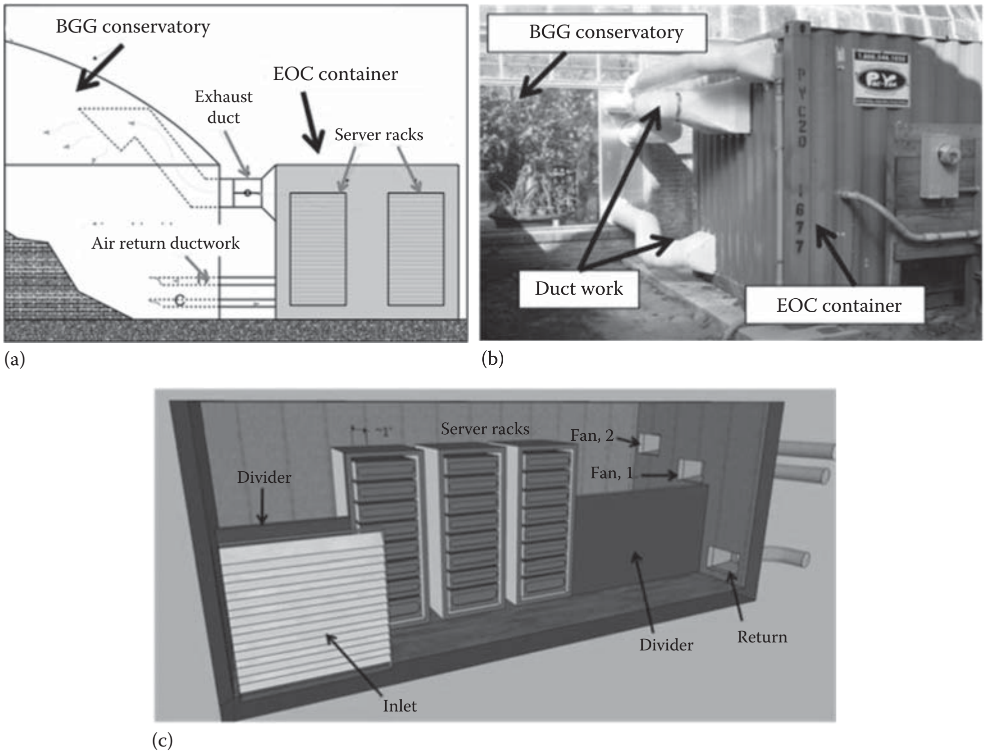


FIGURE 41.1 (a) Layout of prototype EOC container integrated into Greenhouse facility. (b) Photograph of the GC prototype at the Greenhouse. (c) Schematic of prototype EOC container.

recovering waste heat into an existing facility. The University of Notre Dame Center for Research Computing (CRC), the City of South Bend (IN), and the South Bend Botanical Society have collaborated on a prototype building-integrated distributed data center at the South Bend Botanical Conservatory and Greenhouse (Greenhouse) called the Green Cloud (GC) Project [22]. Currently, the ICT infrastructure is based on the University of Notre Dame Condor [23] pool serving a wide variety of high-throughput research computing needs. Jobs can migrate from personal workstations to the traditional data center to dedicated Condor servers at the Greenhouse, where the authors have designed and deployed a sustainable distributed data center (SDDC) container-based prototype. This SDDC serves as a dedicated resource hub for the GC prototype and is shown in Figure 41.1b and c.

The GC hardware resources in the SDDC are fully integrated into the general access Notre Dame Condor pool appearing no different than other resources to end-user jobs. The differentiating factor lies in the environmentally aware controls system set in place for job management and scheduling. The controls AQ6 system currently has two primary components: Condor and xCAT [24]. The Condor component handles the entire workload management and each server's response to the workload based on environment (such as system temperature). The xCAT component handles real-time monitoring of system vitals through interface with the hardware's service processor Intelligent Platform Management Interface (IPMI). The interface between the two is handled with short Python scripts. It is rational to consider a direct interface between Condor and the hardware level diagnostics; however, the robust existing capabilities of xCAT in this regard for a variety of hardware models has made using the two separate software (Condor and xCAT), the most readily viable option. (xCAT is also our standard cluster management tool for server installation and administration.)

The SDDC was designed to minimize capital cost while still providing a suitably secure facility for use outdoors in a publicly accessible venue. The SDDC container is a standard 20 ft long by 8 ft wide shipping container retrofitted with the following additions: a 40 kW capacity power panel with 208 V power supplies to each rack, lighting, internally insulated walls, man and cargo door access, ventilation louvers, small fans, and ductwork connecting the SDDC to the Greenhouse; for a total cost under $20,000. Exterior power infrastructure including the transformer, underground conduit, panel, and meter was coordinated by the local power utility (American Electric Power) and the City of South Bend. The slab foundation was provided by the City of South Bend. The high-bandwidth network connectivity critical to viable scaling is possible via 1 Gb fiber network connectivity to the Notre Dame campus on the St. Joseph County MetroNet backbone.

From the outset, the SDDC was designed for operation utilizing only direct free cooling via outdoor or Greenhouse air; that is, no specialized air conditioning system was integrated in the container. It was also determined that hardware performance and mean time to failure will be evaluated when pushing systems beyond ASHRAE and industry specified limits for temperature and particulate. By allowing the hardware in the SDDC to endure a larger window of thermal fluctuation, we are working to provide variable exhaust heat densities for delivery to the Greenhouse. System level temperatures are provided by the hardware IPMI and validated by occasional infrared camera measurements. The system has multiple inlet, outlet, and fan options, which will allow for hands-on education of mechanical engineering undergraduates studying heat transfer mechanisms.

The eBay Corporation graciously provided over 100 servers for use in this rigorous prototype environment. Per our agreement with eBay, we are not able to provide the reader with specific details on the server models utilized. We can however summarize that they are commercially available multicore systems. The servers began accepting jobs from the Notre Dame Condor pool in December, 2009. Since the SDDC came online, seasonal temperature variations and continued tuning have allowed the numbers of concurrently running jobs to vary from 0 to 250. Dynamic, real-time machine utilization is posted publicly on the GC website. As the afternoon temperatures warm up in the SDDC, machines idle or migrate jobs to keep their core temperatures below specified limits; the jobs return in the evening when the environment is more suitable. In the same temporal cycle, the Greenhouse will become a priority service location in the evening as a power tariff under negotiation with the local utility will provide much lower costs at night.

During moderate-temperature months, external air ($\sim$50°F/10°C) is introduced into the container through a single 54 in. × 48 in. (1.4 m × 1.2 m) louver, heated by the hardware, and expelled into the conservatory. Conversely, during cold-temperature months, when external air is too cold ($<$50°F/10°C) to appreciably heat for benefit to the conservatory, a return vent has been ducted to the conservatory to draw air directly from the conservatory into the container, heat it from the hardware, and then return it directly into the conservatory. Air is driven by a set of three axial fans through two ducts into the Greenhouse. The fans deliver a total volume flow rate of approximately 1260 cfm (3.6 L/min) at a speed of approximately 26.9 ft/s (8.2 m/s) through one duct and 18.4 ft/s (5.6 m/s) through the other. For operation during summer months, when the conservatory does not require additional heating, the ductwork is disconnected and the container simply uses external air cooling for the hardware.

41.4 EOC Control

One of the largest challenges in the development of the GC prototype has been heat management. As opposed to a traditional data center, machines in the GC pool group are additionally managed by an environmentally aware GC Manager (GCM) controls system. This is necessary because the SDDC is not fitted with its own HVAC equipment and relies solely on external air cooling by either the outside air through the louvers or Greenhouse air through a return vent, during the hot and cold seasons, respectively. The primary role of the GCM is to maintain each machine within its safe operating temperatures as stated by their manufacturers by shutting it down if it exceeds them to prevent damage. At the same time, the GCM attempts to maximize the number of machines available for scientific computations, therefore maximizing the temperature of the hot-aisle air that is used for Greenhouse heating.

The GCM interfaces both Condor and xCAT. xCAT provides access, by means of IPMI, to the functionality of the hardware's built-in service processors: power state control and measurements of intake, RAM and CPU temperatures, fan speeds and voltages. The measurements are used by the GCM to decide whether or not the machine is operating within a safe temperature range. The Condor component handles all of the scientific workload management: deploying jobs on running servers, evicting jobs from machines meant to be shut down, and monitoring the work state of each core available in the GC. The GCM posts new Condor configurations to the machines whenever any actions are required.

AQ7

The GCM is written in Python and provides rule-based control of the servers running in the GC based upon xCAT data as well as online measurements of the cold-/hot-aisle temperatures in the SDDC container gathered from the two APC sensors. The rules are applied every 2 min to each machine individually, deciding what action the machine should take under the current conditions: start (machine is started, Condor starts running jobs), suspend (Condor jobs are suspended, machine is running idly), and hibernate (Condor jobs are evicted and migrated to a different machine, machine shuts down). Apart from system control, the GCM also provides detailed logging of the transient conditions. Each log entry is stored separately and consists of air temperature measurements in the SDDC and individual machine values. The logs are saved to provide reference points after adjustments to GCM rules or the physical setup of the prototype are made.

To provide ease in interpretation of the GCM logs, an AJAX-based GC Viewer has been created. This online tool provides a near real-time view of the rack-space with color gradient representation of temperatures and information about core utilization and machine states. It also allows users to choose any data point in the measurement period and run a slideshow-like presentation of the changes to the machine states and temperatures.

As previously mentioned, the SDDC prototype container is fitted with air temperature sensors, whose readings are stored in log files and easily viewed using the GC Viewer. Studying the exact distribution of temperatures in the GC provided us with much needed feedback on how to improve the management logic of the GCM. Experiments covered within this chapter were conducted during the months of June, July, and August 2010. Given that these months are typically the hottest on average year round in South Bend, IN, they represent the worst-case free cooling situation of the GC, which only improve during the remainder of the year. The outside temperatures referenced in the following are based on measurements from weather station MC7428 of Weather Underground, which is nearest to the SDDC prototype. The first trials were designed to run the GC machines within vendor-specified environment temperature ranges to prevent machine wear due to overheating. In the basic rule-set, seen in Algorithm 1, whenever a machine's intake temperature exceeded the stop temperature (106°F) value, it was hibernated. The machine was only restarted after its intake temperature dropped below a starting point (99°F) to prevent excessive power cycling.

Algorithm 41.1 Basic Rule = Set for the GCM

```
if   $mytemp<=$start_temp : start;
if   $mytemp,$sleep_temp : continue;
if   $mytemp>$sleep_temp : hibernate
```

With the basic rule-set, we had 25–30 of 60 machines running Condor jobs during July afternoons with outside temperatures ranging from 82°F to 95°F. Given that the machines were cooled by the ambient air and there was no artificial cooling, this result could be reasonable. However, viewing the situation in GC Viewer provided clues on how this could be improved. As expected, the coolest machines were at the bottom of the racks, and the coolest rack of all was the one placed nearest the intake vent. On all three racks, the top 3–5 machines were always in hibernation due to their excessive intake air temperatures, except for cold early mornings (5:00 a.m.), which effectively rendered them useless. This indicates that the initial physical arrangement of the SDDC air flow is suboptimal for rack cooling using ambient air and could be improved.

Additionally, a very important observation was made. There usually was a large intake temperature difference (8°F–15°F) between machines that were on and running jobs and the ones that were hibernated. The difference was even sharper between machines directly neighboring each other (vertically) in a rack. The issue was because the intake fans do not spin when a machine is hibernated, and therefore, there was insufficient air flow for the machine to cool down below the starting temperature, possibly even causing recirculation of hot air from the hot aisle to the cold one. As a result, during the early hours of the night, there were hibernated machines as hot as 115°F at the top of the racks while machines at the bottom of the racks that were running jobs were registering 75°F intake temperatures.

The previous observation prompted further work on the GCM, which included real-time monitoring of the hot-/cold-aisle temperatures and providing information about the placement of the machine inside the rack. With these being exposed in the rule engine, a new spatially aware rule-set, shown in Algorithm 2, was introduced. The rules regarding hibernation were split into two, depending on the temperature of the cold aisle. When the cold aisle was hot (above 85°F), the behavior stayed the same as in the basic approach. In colder cases, the behavior changed. Rules would suspend the machine from performing Condor jobs instead of hibernating it after exceeding the sleeping temperature (106°F). This would keep the internal server fans spinning, facilitating more rapid cooling of the machine below the starting temperature (99°F). The machine would only be hibernated (turned off) if it exceeded an upper operating temperature (109°F). This rule prevented newly started machines from rehibernating more quickly than they could cool down.

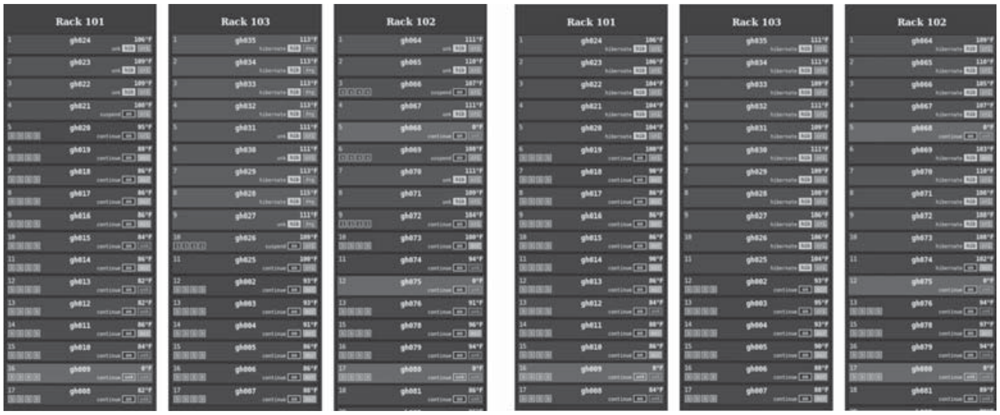
The biggest difference came because of the introduction of the last rule. It would force hibernated machines to wake up, even if they reported modest overheating, in order for the fans to spin up and cool the machine down. However, this behavior was only applied if the average intake temperature of the

Algorithm 41.2 Spatially Aware Rule-Set for the GCM

```
if   $mytem<=$start_temp : start;
if   $mytemp<$sleep_temp : continue;
if   $cold_aisle>85 and $mytemp>temp=
     $sleep_temp : hibernate;

if   $scold_aisle <=85 and $mytemp>=
     $sleep_temp and $mystate= =”on”
     :suspend;
if   $scold_aisle <=85 and $mytemp>=
     $danter_temp : hibernate;

if   $cold_aisle <=85 and $mystate!=”on”
     and $mytemp<=$danter+temp+6 and
     $neightemp<=$start+temp : start;
```



(a) (b)

FIGURE 41.2 Comparison of rule-set data points in GC Viewer. Note the white/black boxes representing hibernation/running power states, respectively. Only the top 17 machines are shown for each rack, due to space limitations. (a) [left three columns] Spatially aware rule-set (August 1, 2010, 20:00 p.m.): running: 41, hibernated: 19, cold aisle: 82°F, hot aisle: 111°F. (b) [right three columns] Basic rule-set (August 2, 2010, 20:00 p.m.): running: 33, hibernated: 27, cold aisle: 84 F, hot aisle: 109°F.

AQ8

two machines placed directly beneath it was below the starting point temperature (99°F). This prevented rapid reheating due to natural convection from its nearest neighbors.

The impact of the spatially aware rule-set can be seen in Figure 41.2. During two consecutive days of similar weather, we ran both rule-sets and measured their impact on the state of the GC during the evening (8 p.m., 81°F outside temperature in both cases). While using the basic rule-set, only 33 machines were running and 27 were hibernated. Using the spatially aware rule-set, the number of running machines rose to 41, meaning 36 more processor cores were available to run computations. Moreover, in the case of the spatially aware rule-set, the hot-/cold-aisle temperature difference rose as well, resulting in a 13.7% increase, from 9.72 to 11.29 kW, in available waste heat for the Greenhouse. We observed that even during the hottest parts of the day, we could run 2–3 more machines using the spatially aware rule-set.

While the controls algorithm are important for the reality of suboptimal container and environmental conditions, it should be noted that the availability of outside air free cooling (below 95°) has the potential to provide nearly year round server operation (no suspension, etc.).

41.5 EOC Thermal Measurements

Thermal measurements were conducted on the SDDC prototype during the summer months of June, July, and August of 2010 when the container was fully operational, but not actively heating the conservatory. The thermal measurements consisted of the constant monitoring of various local air temperatures throughout the container as well as server temperatures and server loads. In this way, local temperature and heat recovery could be estimated and directly correlated to server usage and activity. The prototype container was analyzed as a single, pseudo-steady state control volume as shown in Figure 41.3 where q_{waste} is the heat generated by the servers (as well as other electrical equipment), q_{in} is the energy advected into the container through the louvers, q_{loss} accounts for any thermal loss through the insulated container walls and doorways, and q_{out} is the heat leaving the container to effectively heat the Greenhouse.

In principle, the waste heat delivered to the Greenhouse is simply the difference between the exhausted heat q_{out} and the incoming heat not generated by the servers q_{in}. Because the configuration of the servers

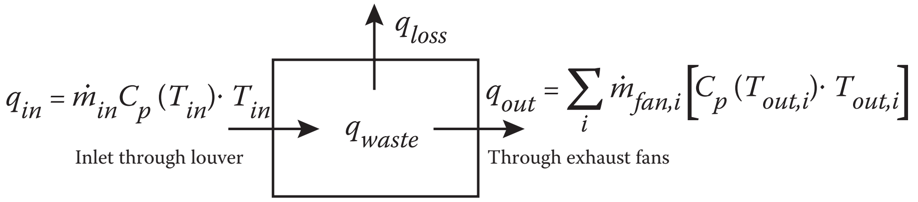


FIGURE 41.3 Schematic of heat flow through the prototype container.

and hot/cold aisles was not optimal, the true waste heat delivered could not be easily measured. Therefore, for this preliminary study, the amount of waste heat available for recovery was estimated as the amount of heat picked up across the servers minus any heat losses.

$$q_{waste} = \dot{m} \left[c_p(T_{ha}) T_{ha} - c_p(T_{ca}) T_{ca} \right] - q_{loss}, \tag{41.1}$$

where

$c_p(T)$ is the specific heat of the air at the local temperature

T_{ca} and T_{ha} are the temperature of the cold-aisle upstream of the servers and the hot-aisle downstream of the servers, respectively

It is difficult to exactly quantify the heat loss q_{loss} through the insulated container walls, doorways, and louver. For this preliminary analysis of the prototype data, the heat loss was estimated using a 1D conduction analysis through the insulated walls with thermal conductivity of 0.0058 Btu/(h ft°F) (0.01 W/m K), by assuming the wall exteriors were at ambient and the wall interiors were either at the hot-aisle or cold-aisle temperatures. The mass flow rate was determined by applying mass conservation and calculating the total flow rate passing through the two exhaust fans:

$$\dot{m} = \sum_{i=1,2} \dot{m}_{fan,i} = \sum_{i=1,2} \rho \left(T_{out,i} \right) U_{avg,i} A_{duct,i}, \tag{41.2}$$

where

i indicates the two exhaust fan ducts

ρ is the local air density

A_{duct} is the cross-sectional area of the exhaust duct for each fan

$U_{avg,i}$ is the average flow speed at the fan exit

The exit flow speed was measured one time (the fans operated at constant speed) using a handheld hot wire anemometer with integrated thermocouple (Extech Model 407123) with an accuracy of approximately $\pm 3.0\% + 0.3$ m/s. The velocimeter was placed at the exit of the fan duct, perpendicular to the flow, to record flow speeds for 20 s, and the time-averaged speed was calculated. To obtain the average exit flow speed, the same measurement was conducted over a number of locations across the duct exit (Figure 41.4), and the average flow speed was calculated as

$$U_{avg} = \frac{1}{A_{duct}} \int_{A_{duct}} U(r, \theta)\, r dr d\theta. \tag{41.3}$$

The integration was conducted numerically using trapezoidal rule, and the uncertainty in the mass flow rate was estimated to be ±8.1%.

Air temperatures in the SDDC were recorded with four temperature/humidity sensors (APC Model AP9512THBLK) connected to two networked control boxes (APC Model AP9319). One sensor was placed just upstream of each exhaust fan ($T_{out,i}$), one sensor was placed in the cool aisle (T_{ca}), and one

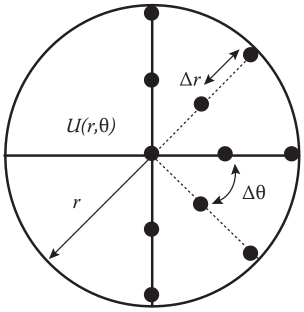


FIGURE 41.4 Schematic of measurements to determine the average outlet velocity. The radii of ducts 1 and 2 were $r = 5$ in. and $r = 4$ in., respectively. The spacing between measurements was $\Delta\theta = \pi/4$ and $\Delta r = 2.5$ in. for duct 1 and $\Delta r = 2$ in. for duct 2. Symmetry was assumed and the flow was only measured in two quadrants.

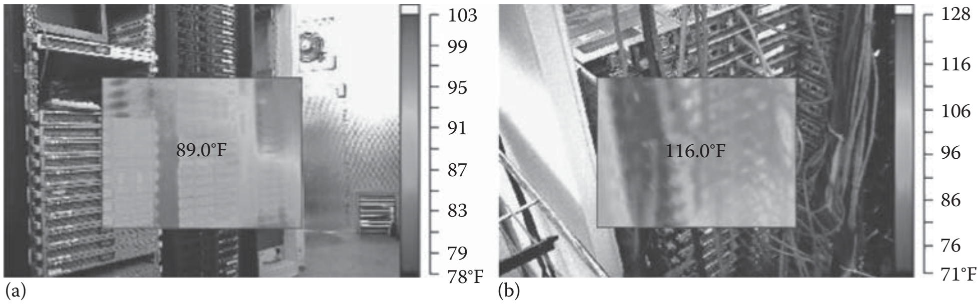


FIGURE 41.5 Illustrative infrared thermal maps of the server (a) inlet and (b) outlet temperatures.

was placed in the hot aisle (T_{ha}). Temperatures were recorded in real time at a rate of four readings per hour. At present, the prototype is not configured to enable temporal temperature measurements in the ducts downstream of the fans, which would be the best way to accurately measure heat recovery. The temperature at the inlet of the louver T_{in} was taken from weather measurements from the Indiana State Climate Office (ISCO 2010). The heat recovery as a function of time $q_{rec}(t)$ was estimated using Equation 41.1 along with the sensor readings, and mass flow rate was estimated using Equation 41.2. In addition to occasional spot checks with a handheld thermocouple for assurance that the sensors were correct, the temperature readings were also validated by comparisons with the temperatures of the HPC hardware as well as occasional infrared camera measurements (Figure 41.5). Hardware temperatures (T_{HPC}) were recorded from the hardware's internal temperature sensors using the IPMI.

Figure 41.6a shows a representative plot of the temperature measurements from the four sensors, the estimated fan downstream temperatures, the inlet temperature, and the temperatures from one of the servers over a 48 h period. The plot illustrates three significant points. The temperature of the hot aisle is significantly hotter than that of the cold aisle, exemplifying the vast amount of waste heat that is generated in data centers, and, though the single hardware temperature (T_{HPC}) varies significantly because of dynamic computational loads, overall the temperatures are fairly constant because the total number of active servers stays fairly constant. Finally, the average server inlet temperatures range from $\sim$ 70°F to 100°F (21°C–38°C), which exceeds current recommended HPC hardware operating ranges. One aspect of the EOC philosophy is that ITC can be operated beyond current limits, and the data demonstrates server operation at temperatures greatly exceeding standards. AQ9

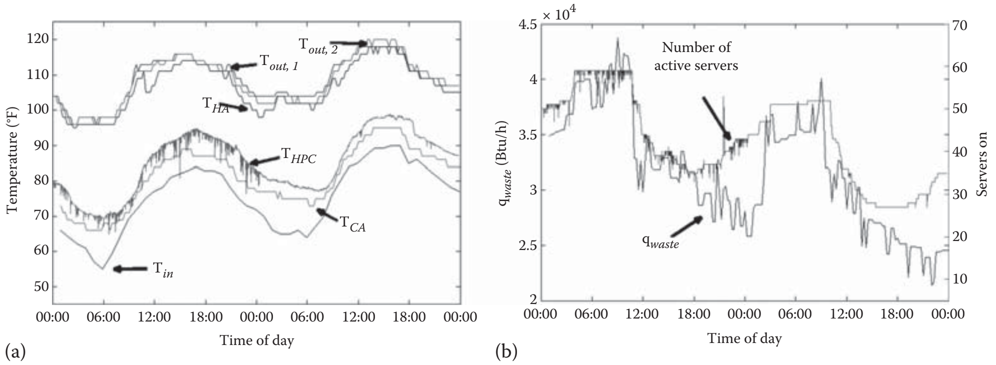


FIGURE 41.6 (a) Representative measured temperatures and (b) available waste heat over a 48 h period in July 2010.

Figure 41.6b shows the amount of waste heat available for recovery for the same 48 h period. On average, nearly 33.6×10^3 Btu/h (9.39 kW) was extracted from the data servers for this period of time, for a total energy recovery of 1.61×10^6 Btu (450.7 kW h). Though this value is limited by approximations for the heat loss and simple bulk temperature measurements, it was consistent with the energy consumed by the container according to energy bills during the same period. At an estimated cost of \$0.10/kW h, the measured heat recovery corresponds to a \$45.07 energy savings for this time period, and, when extrapolated to a month, approximately \$676, or ∼4.5% of the average monthly expenditures for the Greenhouse during a winter month (an average ∼\$15, 000 during the months of December through February from 2003 to 2006). In these preliminary studies, the container was limited to a maximum of 60 servers operating at any time in part due to the operational layout but primarily because the warm summer air necessarily increased T_{ca}. With improved configuration and in cooler months, this prototype container should be able to operate at a nearly constant heat recovery of 102,364 Btu (30 kW) corresponding to ∼15% savings in the Greenhouse's monthly energy consumption during the winter. Further, the installation of additional containers will provide incremental increases to the total savings accordingly.

One of the primary challenges of successfully integrating HPC into built structures is balancing the market forces illustrated previously in Table 41.1. Successful energy recovery from an EOC container requires that heat be delivered to its partner facility when it is required and in a predictable manner. With EOC, the availability of heat depends on the computing users' need for computational power at that given moment. This basic issue is illustrated in Figure 41.7 when a preset pilot test was conducted during which the computational load on the servers was intentionally varied intermittently—the servers were initially idle for a 27 h period followed by a period of normal loading capacity for 12 h, idle for 12 h, active again for 12 h, and then idle for 12 h. As the pilot test demonstrated, the temperature rise not only reduces dramatically when the HPC hardware is idle, but there is also a transient recovery period when the hardware is active but the temperature rise follows more slowly. The thermal time constant of this system is related to not only the heat capacity of the air, but of the entire set of server components and infrastructure (racks, etc.) as well, which serve as heat sinks during any heat up time. For this prototype, the time constant, when the temperature rise reached 95% of maximum, was estimated to be 53 min. Similarly, during times when the servers were idle, the EOC container took approximately 44 min to reach its ambient condition. In an application where EOC containers are distributed across multiple facilities, control algorithms will be needed to balance the demands of both the compute and building users to balance performance for each set of customers, and such time constants will play an integral role.

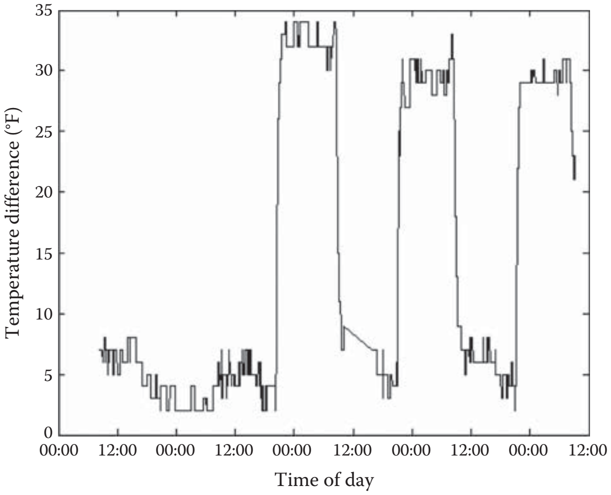


FIGURE 41.7 Temperature measurements for a pilot test over a 72 h period where the servers were left inactive for a 27 h period followed by the servers being operated and then idled for alternating 12 h periods.

41.6 Related Work

It is important to note the evolving commercial applications in grid, utility, and cloud computing [25,26] that will directly benefit from this technology. As the computational infrastructure configuration and locality are obscured from the end user, the flexibility to distribute and configure grows, allowing for additional economic and environmental optimizations. Along these lines, the growing acceptance of virtualization [27,28] in commercial applications will also allow greater flexibility in the design and deployment of EOC-based solutions.

While these grid frameworks are evolving, a large body of work has studied the problem of managing energy, heat, and load in large centralized data centers. Schmidt et al. [29,30] provide a good overview of the mechanical issues of cooling units, heat sinks, fans, and so forth. Server management techniques can also be applied to reduce energy costs. For example, inactive servers can be shut down, or loads migrated as more energy efficient hardware become idle/available. Chase [31] and Bradley [32] describe techniques for balancing performance, cost, and energy in this situation. To avoid hot spots, it is necessary to map the relation between components and heat [33], and then shape loads so as to evenly distribute the heat. AQ10 Further, large institutions such as the University of Illinois and NCSA are taking a holistic look at their entire campus utilities infrastructure to efficiently operate their data center. Despite the new efficiency benefits, Patel et al. [34] report that a typical data center still consumes about as much energy for cooling as it does for productive work. The advances toward seven more efficient traditional infrastructures and EOC frameworks will serve in tandem to provide greater computational capability while reducing economic and environmental costs.

41.7 Conclusions

In this work, we have introduced the principles of EOC and demonstrated its efficacy with the GC prototype that is integrated with a City of South Bend botanical garden and greenhouse. The prototype successfully conducts significant amounts of workload from the University of Notre Dame Condor pool despite geographical separation and operating only with outside air cooling. It has become a viable tool for analysis and research into the management of ICT energy reutilization, server states, and load distribution. To that end, EOC provides economic and environmental benefits to improve the sustainability of information and communication technology infrastructure.

Our GC prototype serves as a successful demonstration that ICT resources can be integrated with the dominate energy footprint of existing facilities and dynamically controlled to balance process throughput, thermal energy transfer, and available cooling via process management and migration. Apart from not using energy-consuming air conditioning cooling at a centralized data center, the GC further improves energy efficiency by harvesting the waste hot air coming from the working servers for the adjacent Greenhouse facility. The success of this technique makes the EOC concept even more attractive for sustainable cloud computing frameworks.

EOC is a sustainable computing technology that complements existing efficiency improvements at the application, operating system and hardware levels. With the growing enterprise utilization of cloud computing and virtualized services, EOC becomes more viable in across the range of ICT services. The measured average and maximum throughput continues to improve as we tune our GCM controls system and modify the physical layout of the prototype. This initial success will stimulate continued project growth to the economic and environmental benefit of both our organization and our community partners.

Acknowledgments

The authors would like to recognize contributions to this work from multiple organizations and individuals. The City of South Bend and South Bend Botanical Society have been close partners contributing funds and facility expertise. The eBay Corporation generously donated over 100 servers to the Green Cloud prototype. Dr. Jaroslaw (Jarek) Nabrzyski and the Center for Research Computing operations group contributed essential financial resources and system administration expertise. Our undergraduate and summer interns Ryan Jansen, Eric Ward, and Michal Witkowski were key contributors to the infrastructure development and operations. Partial support for this work was provided by a Department of Energy grant for the Northwest Indiana Computational Grid.

References

1. U.S. EPA. August 2007. Report to congress on server and data center energy efficiency public law 109–431. Technical Report. Washington, DC: U.S. Environmental Protection Agency. AQ11
2. Blazek, M., Mills, E., Naughton, W., Tschudi, P., Sartor, D., Seese, R., and Shamshoian, G. 2007. The business case for energy management in high-tech industries. *Journal of Energy Efficiency*, 1:1–16.
3. ASHRAE Technical Committee 9.9. 2008a. Best practices for datacom facility energy efficiency. Atlanta, GA: American Society of Heating Refrigeration, and Air-Conditioning Engineers.
4. ASHRAE Technical Committee 9.9. 2008b. High density data centers—Case studies and best practices. Atlanta, GA: American Society of Heating Refrigeration, and Air-Conditioning Engineers.
5. Brill, K.G. 2008. Special report: Energy efficiency strategies survey results. Technical Report. Santa Clara, CA: The Uptime Institute.
6. AQ12 Green Grid: 7 × 24 Exchange International, ASHRAE, Silicon Valley Leadership Group, U.S. DOE Save Energy Now Program, Energy Star Program: U.S. EPA, U.S. GBC, Uptime Institute. 2010. Recommendations for measuring and reporting overall data center efficiency. Version 1: Measuring PUE at Dedicated Data Centers.
7. Barcelona Supercomputing Center—Centro Nacional Supercomputación. 2006. Annual Report 2005. Barcelona, Spain.
8. AQ13 Fontecchio, M. 2008. Data center news: Companies reuse data center waste heat to improve energy efficiency. Winnipeg, Manitoba, Canada: Search Data Center.

9. Alger, D. 2010. *Grow a Greener Data Center*. Indianapolis, IN: Cisco Press.
10. Miller, R. February 2010. *Data Centers Heat Offices, Greenhouses, Pools*. Lawrenceville, NJ: Data Center Knowledge.
11. AQ14 Ahmad, I., Ranka, S., and Khan, S.U. Using game theory for scheduling tasks on multi-core processors for simultaneous optimization of performance and energy. In *The Next Generation Software (NGS) Workshop 2008*, Melbourne, Victoria, Australia, April 2008.
12. Gunaratne, C., Christensen, K., and Nordman, B. 2005. Managing energy consumption costs in desktop PCs and LAN switches with proxying, split TCP connections, and scaling of link speed. *International Journal of Network Management*, 15(5):297–310.
13. Jejurikar, R., and Gupta, R. Dynamic voltage scaling for systemwide energy minimization in real-time embedded systems. In *Proceedings of the 2004 International Symposium on Low Power Electronics and Design (ISLPED'04)*, Newport, CA, 2004.
14. Sharma, S., Hsu, C.H., and Feng, W.-C. Making a case for a green500 list. In *IEEE International Parallel and Distributed Processing Symposium 2006, Workshop on High Performance Power Aware Computing*, Rhodes Island, Greece, 2006.
15. Eilam, T., Appleby, K., Breh, J., Breiter, G., Daur, H., Fakhouri, S.A., Hunt, G.D.H., Lu, T., Miller, S.D., Mummert, L.B., Pershing, J.A., and Wagner, H. 2004. Using a utility computing framework to develop utility systems. *IBM Systems Journal*, 43:97–120.
16. Figueiredo, R., Dinda, P., and Fortes, J. 2005. Resource virtualization renaissance. *Computer*, 38:28–31.
17. ASHRAE. 2004. *ANSI/ ASHRAE Standard 55-2004, Thermal Comfort Conditions for Human Occupancy*. Atlanta: American Society of Heating, Air-Conditioning and Refrigeration Engineers, Inc.
18. USGBC. 2007. *LEED for New Construction & Major Renovation Reference Guide,* Version 2.2. Indoor Environmental Quality: EQ Credit 6.2, Controllability of Systems: Thermal Comfort. Washington, DC: USGBC.
19. Brenner, P., Thain, D., and Latimer, D. Grid heating clusters: Transforming cooling constraints into thermal benefits. In *Uptime Institute—IT Lean, Clean, and Green Symposium*, New York, 2009.
20. Brenner, P., Thain, D., and Latimer, D. April 2008. Grid heating: Transforming cooling constraints into thermal benefits. Technical Report 2008–30, Computer Science and Engineering Department, University of Notre Dame, Notre Dame, IN.
21. Lammie, M., Brenner, P., and Thain, D. Scheduling grid workloads on multicore clusters to minimize energy and maximize performance. In *10th IEEE/ACM International Conference on Grid Computing,* Banff, Alberta, Canada, 2009.
22. Brenner, P., Jansen, R., Go, D., and Thain, D. Environmentally opportunistic computing transforming the data center for economic and environmental sustainability. In *First International Green Computing Conference, Technically Co-Sponsored by IEEE Computer Society*, Chicago, FL, 2010.
23. Litzkow, M., Livny, M., and Mutka, M. Condor—A hunter of idle workstations. In *Eighth International Conference of Distributed Computing Systems*, San Jose, CA, June 1988.
24. Ford, E. 1999. xCAT extreme cloud administration toolkit. IBM internal software development. Open Sourced 2007.
25. Raghavan, B., Vishwanath, K., Ramabhadran, S., Yocum, K., and Snoeren, A.C. 2007. Cloud control with distributed rate limiting. In *SIGCOMM'07: Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, Kyoto, Japan, pp. 337–348, New York: ACM.
26. Mitchell, W. August 2007. Data center in a BOX. *Scientific American*, 297:90–93.
27. Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I., and Warfield, A. 2003. Xen and the art of virtualization. In *SOSP'03: Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles*, Bolton Landing, NY, pp. 164–177, New York: ACM.

28. Uhlig, R., Neiger, G., Rodgers, D., Santoni, A., Martins, F., Anderson, A., Bennett, S., Kagi, A., Leung, F., and Smith, L. 2005. Intel virtualization technology. *Computer*, 38:48–56.
29. Schmidt, R., Cruz, E., and Iyengar, M. July 2005. Challenges of data center thermal management. *IBM Journal of Research and Development*, 49(4–5):709–724.
30. Schmidt, R., and Iyengar, M. January 2007. Best practices for data center thermal and energy management—Review of literature. *ASHRAE Transactions*, 113(1):206.
31. Chase, J., Anderson, D., Thakar, P., Vahdat, A., and Doyle, R. Managing energy and server resources in hosting centers. In *Symposium on Operating Systems Principles*, Banff, Alberta, Canada, 2001.
32. Bradley, D., Harper, R., and Hunter, S. 2003. Workload based power management for parallel computer systems. *IBM Journal of Research and Development*, 47:703–718.
33. Kang, J., and Ranka, S. Assignment algorithm for energy minimization on parallel machines. In *2009 International Conference on Parallel Processing Workshops*, Vienna, Austria, 2009.
34. Patel, C.D., Bash, C.E., Sharma, R., and Beitelmal, M. Smart cooling of data centers. In *Proceedings of IPACK*, Maui, Hawaii, July 2003.

AUTHOR QUERIES

[AQ1] Please provide affiliation details for Douglas Thain, Aimee Buccellato, and David B. Go.
[AQ2] Please provide expansion of "DOE" if appropriate.
[AQ3] Please provide the expansion of "EPA" if appropriate.
[AQ4] Please check if the edit to the sentence beginning "Comfort system control . . ." is ok.
[AQ5] Please note that abbreviation "GC" is used for both "grid heating" and "green cloud" in the chapter. Please check.
[AQ6] Please provide expansion of "xCAT," if appropriate.
[AQ7] Please provide expansion of "APC" if appropriate.
[AQ8] Please check if the edits to figure caption 41.2 are ok.
[AQ9] Please provide the expansion of the acronym "ITC," if appropriate.
[AQ10] Please provide expansion of "NCSA" if appropriate.
[AQ11] Please provide location details for Refs. [1,5].
[AQ12] Please provide the Publisher location for the Ref. [6]
[AQ13] Please check this updations in the Ref. [8,10,29,30], for correctness.
[AQ14] Please provide location details and proceeding/conference date for the following references: 11, 13, 14, 19, 21, 22 23, 31, 33, and 34.